R E P O R T R E S U M E S

ED 013 634 JC 670 854

THE AUTO-CRITIQUE METHOD OF INSTRUCTIONAL EVALUATION.

BY- ANDERSON, JOHN E.

PUB DATE JUL 64

EDRS PRICE MF-$0.25 HC-$0.68 17P.

DESCRIPTORS- *JUNIOR COLLEGES, *TEACHER EVALUATION, *SELF EVALUATION, TEACHER IMPROVEMENT, INSTRUCTIONAL IMPROVEMENT, COLUMBUS COLLEGE, COLUMBUS, GEORGIA,

OF 26 FULL-TIME MEMBERS AT COLUMBUS COLLEGE, 19 VOLUNTARILY PARTICIPATED IN A STUDY OF A SELF-EVALUATION PROCEDURE. USING A 7-POINT SCALE, EACH PARTICIPANT RATED HIMSELF ON (1) SPEAKING VOICE, (2) MANNERISMS, (3) KNOWLEDGE OF SUBJECT, (4) HIS ENTHUSIASM, (5) CLASS ENTHUSIASM, (6) DIGRESSIONS, (7) ORGANIZATION AND PREPARATION, (8) USE OF ANALOGIES, EXAMPLES, AND ILLUSTRATIONS, (9) HANDLING OF QUESTIONS, AND (10) GENERAL CLASS ATMOSPHERE. HE THEN RECORDED TWO CLASS SESSIONS DURING A 2-WEEK PERIOD, AND FOLLOWED HIS REVIEW OF THE TAPES BY ANOTHER SELF-RATING. ALTHOUGH NO SIGNIFICANT DIFFERENCES WERE FOUND IN THE TWO RATINGS, AFTER LISTENING TO THE TAPES OF THEIR CLASSES, FIVE TEACHERS RATED THEMSELVES MORE FAVORABLY AND SIX PLACED THEMSELVES LOWER ON THE SCALES. THUS, THERE IS EVIDENCE THAT OVER HALF OF THE PARTICIPANTS WERE SENSITIVE TO THE INFORMATION OBTAINED FROM THE TAPES. THE AUTHOR CONCLUDED THAT, ALTHOUGH LACK OF EXTERNAL CRITERIA MAKES THIS PROCEDURE UNSUITABLE FOR MERIT RATING PURPOSES OR FOR COMPARISON OF ONE INSTRUCTOR WITH ANOTHER, THE METHOD HAS PROMISE FOR THE INDIVIDUAL TEACHER'S SELF-IMPROVEMENT ACTIVITIES. (WO)

From: Improvement of Instruction, Proceedings of the Fourth Junior College Administrative Teams Institute, July 27-31, 1964, Pensacola, Florida.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

UNIVERSITY OF CALIF.
LOS ANGELES

OCT 3 1967

CLEARINGHOUSE FOR
JUNIOR COLLEGE
INFORMATION

ED013634

THE AUTO-CRITIQUE METHOD OF INSTRUCTIONAL EVALUATION

John E. Anderson, Dean
Columbus College

As the population explosion continues to be felt in the colleges and universities, the expansion of present facilities and the planning of new ones have become part of the educational scene on campuses all across the country.

As new buildings and colleges spring up as if by magic, legislators and the public may point with pride at the tangible evidence of their concern about the future of higher education in America. But, the creation of each new campus, the completion of each new building, and even the establishment of each new curriculum are not without their drawbacks. For the problem of staffing each facility is no longer a bridge to be crossed in the future, it is imminent, and the available supply of experienced faculty is limited. In fact, at present, most institutions have one or more unfilled vacancies on their staffs because of the lack of qualified candidates. Further, data from graduate schools and other sources of potential faculty indicate that the shortage will become increasingly severe into the 1970's. Consequently, department heads, deans, and presidents have intensified their efforts to not only acquire but retain competent qualified faculty for their institutions.

JC670854

To this end, armed with data concerning soaring enrollment, teacher shortage, and current competitive salaries, most presidents have been able to slowly but surely extract increasing amounts of money from legislators or trustees for faculty salary raises. This has been a continuing process and promises to be more so in the foreseeable future. However, this largesse is not without its attached strings. Insistence has been made that raises be made contingent upon merit--and not given across the board. The dean, unenviable creature that he is, when confronted by the president on this point, agrees wholeheartedly and with rapidity, because he is under pressure to cover classes and the projected salary increase may spell the difference between a few restful nights and the nightmare of searching for two or three newly vacated positions. Thus assurances are given that only the meritorious shall be given increments.

Of course this entire situation would be rather amusing if it weren't so tragic--and so real! Because for years the brightest and best trained men have gone into industry and more recently into research as the salary structure there has disproportionately increased. Now, after years of selective starvation, the college personnel are asked to reward only the outstanding, and mirabile dictu, there wtill remain in colleges, some who are outstanding. This is either a tribute to the dedication and altruism of the teacher, or a monument to his colossal

ineptitude.

Be that as it may, however, most components of the educational system are happy. The faculty has its raises, the public has its newspaper account of how merit raises only are to be given, the legislature or trustee body has the promise of the board of control or regents, the regents have the promise of the presidents, the presidents have the assurance of their deans, and finally, the deans have their classes covered and ulcers palliated temporarily--but only temporarily.

But soon, as swiftly the dawn the night, comes the day of reckoning. Someone, typically the dean, must show evidence of good faith in attempting to either devise or discover a discriminating merit scale, or else be prepared to defend his conclusion that all his faculty are indeed meritorious, thus deserving of an increment across the board. Since we usually end up with the latter approach anyway, the scope of this paper is limited to the first alternative.

The most frequent and most discussed approach to merit rating is the construction of a scale or index which would reveal good or desirable pedagogical practices. Most have been oriented around the concept of good teaching. Rating scales or other evaluative devices, filled out by students, former students, peers, or superordinates, have been most prevalent. However, the many reliability and validity problems, including definitions and criterion

measures, in addition to the attendant faculty morale disturbances, will not be dealt with since they are well known to all administrators. Generally, this approach has been attempted, discarded, attempted again with a new twist, and subsequently discarded again almost with a monotonous regularity. The question of whether this technique is better than nothing remains moot in the eyes of many deans.

Another approach, less formal and less informative also, is class visitation. Often this is done by the dean, department head, a seasoned and respected teacher, or a peer. Problems of open faculty resentment and sampling bias are of course manifest, but of more pragmatic concern to the dean is the inordinate amount of time which must be devoted by the visiting evaluators to the visitation schedule. Since, if any type of systematic and comparative information is to be derived, a definite plan and adequate coverage must be obtained, but, because of the time factor that this arrangement necessarily entials, most deans have not relied heavily on this method of assessment.

A third approach has been the follow-up of former students in subsequent relevant courses. At the junior college this has been particularly difficult because of the necessary cooperation of the many and diverse senior institutions to which transfers have been sent. In addition, because of the time delay involved, faculty are frequently gone by the time such data are available. Senior

institutions have somewhat the same problems with reference to graduate schools, but within their four-year confines, institutional research programs could be most revealing in this area. To my knowledge, there have been no published data on this to date. This method has not been used to its fullest extent.

A fourth approach to the assessment of faculty merit, and one heavily leaned upon, is that of talks, lectures, symposia, or publications. But, that there is a considerable difference between quotidian class delivery and special discursive forums or written articles, is too obvious to dwell upon. The chief positive values of this approach are that it is quantifiable, gets good press, and, of course, is easy to do. The chief drawback is that it is not relevant.

A fifth approach is rather--and better said--no approach at all, but a melange of rumor, student reports, hearsay, and the typical scuttlebutt pervading a college campus. As a feedback mechanism, this approach has heuristic value, but as a discriminandum for merit it is not only useless, but can do a great deal to destroy faculty morale, and the respect of the dean by the faculty.

A sixth approach, and the one on which this paper reports, is that which I term the auto-critique method. The basic ingredient is that the instructor evaluates himself. The chief disadvantages are that the method is essentially self contained, is not readily relatable to an external criterion, and is not amended to tangible control or

manipulation by the administration. But, since upon scrutiny, the other approaches have these two disadvantages as well, as has been pointed out previously, the advantages of this method offer themselves as worthy of the consideration of a dean who is attempting to discharge his obligation to assess and upgrade the instructional level of his institution. Amond these advantages are:

1. Evidencing interestest in the teaching process itself by the administration,

2. Indicating confidence by the administration in the faculty's ability to evaluate themselves as professionals and make self indicated improvements,

3. Giving the faculty a workable and frequently interesting method of whereby they may improve themselves,

4. Preservation of anonymity by faculty, thus forestalling feelings of "big brother" watching,

5. Establishing essentially a self operating and perpetuating system not calling for a great amount of time,

6. Placing of the dean in the position of being called in for aid by a motivated faculty member, rather than being looked upon as an intruder with unwanted advice,

7. Providing specific and concrete examples (preserved on tape) of problem areas which can be referred to upon replay, and not having to rely on notes or faulty memory.

Method

The materials used in the current investigation were a rating sheet and a tape recorder with two on-hour tapes. Nineteen members of the 26 full-time faculty volunteered to participate. The rating scale was constructed after consultation with various experienced former faculty colleagues of the author at Florida State University, and a review of the items appearing in various published teacher

rating scales. Ten different areas were covered with the categories of "verbal delivery and "enthusiasm" being further subdivided. A seven point rating scale was then imposed on each category with the directions asking the instructor to rate himself from low (1) to high (7) as to his performance in each category. The final category items were:

1. Speaking voice
2. Mannerisms or pleonasms
3. Knowledge of subject matter
4. Enthusiasm - self
5. Enthusiasm - engendered in class
6. Digressions
7. Organization and preparation
8. Use of analogies, examples, illustrations, etc.
9. Handling of questions
10. General atmosphere created.

Anonymity was preserved, although for statistical treatment each instructor was asked to assign himself an easily remembered code name or number. After the form was filled out, it was then turned into the dean's office. Two one-hour tapes and a recorder were then made available to each instructor. His instructions were to start the tape at the beginning of the class hours, and turn it off at the end. This was to be done for two class periods within a two-week period. After both recordings were made, the instructor was to pick up another copy of the rating

sheet, listen to both tapes, and then refill out the rating sheet which was also turned into the dean's office. Thus a comparison could be made of the instructor's original and subsequent ratings of himself. Tapes were reused at will by the instructor. Instructors were encouraged to discuss specific problems emerging from the sessions with an experienced teacher, the department heads, or the dean.

The forms were then analyzed for information.

Results

Since the ultimate purpose and hence the real value, of this auto-critique method was to have each instructor devote his professional skills, talent, time, and energy to a critical self assessment, statistical treatment of group data was considered of subordinate value. Nevertheless, cursory analyses were made and certain conclusions reached. These data will be presented below. Comments elicited from the faculty reflecting the self assessment component of the method are also listed. Thus, both quantitative and qualitative aspects are noted.

The data presented in Table 1 are by columns:

Column 1. The pooled mean score (N=19) by category on the first or Before administration of the scale.

Column 2. The same for the 2 or After.

Column 3. Directionality of change.

Column 4. Pooled rank of each category in comparison with the other categories on the Before.

Column 5. Same on After.

Below the table you will note the grand mean for the

Table 1

	Mean Score Before	Mean Score After	Direction of change of Mean Score	Rank Before	Rank After
1.	5.64	5.72	+	6	9
2.	4.50	4.26	-	10	8
3.	6.16	6.15	-	1	1
4.	6.05	5.89	-	3	2.5
5.	4.89	5.21	+	9	10
6.	5.37	5.73	+	8	6
7.	5.71	5.78	+	5	3.5
8.	6.11	5.89	-	2	2.5
9.	5.79	5.78	-	4	3.5
10.	5.47	5.52	+	7	7
Grand Mean	5.56	5.63	+ rho = .87 sig. .01		

Table 1 Mean scores, direction of change, and ranks on Before and After administrations of rating scale.

Before and the After administrations, and a Spearman rank order correlation coefficient.

The statistical analysis included:

1. An analysis of variance and F test.

2. White's Rank Order test of significance.

3. Spearman Rank Order correlation method and test of significance for sampling variance.

As can be readily inferred, no statistically significant differences between the Before and After administrations of the questionnaire were found using either the parametric analysis of variance or the non parametric Rank Test. Further, the coefficient of correlation between the two administrations of the scale was .87, significant beyond the .01 level, hence indicative of a quite high and reliable relationship. The conclusion that there was not an overall change in self rating on this seven point scale as a function of the listening to the taped sessions is readily made. However, since on a rating scale in which there is a compressed continuum, i.e., seven points plus a considerable halo effect operating--note the skewing toward the high rather than middle or low end of the scale--a factor suggesting that the faculty think highly of themselves on both administrations, which is not unrealistic since the college thinks highly of them also, and has reinforced this view both with verbal praise and behavioral acts. On such a rating scale, it is not unexpected that numerical data are not significant. However, much information is available from inspective analysis of the pattern

of responses made. Along these lines, some interesting points emerge. They are:

1. Widest range = Category 5, (Enth-Eng-Class)
2. Constricted = Category 9,(Handling Questions)
3. Highest rank Before=Category 3, (K of Sugj.Matter)
4. Highest rank After =Category 3, (K of Subj.Matter)
5. Lowest rank Before =Category 2, (Mannerisms or Pleonasms)
6. Lowest rank After =Category 5, (Enthus.-Class)
7. Of special interest is Category 6, (Digressions), which shifted from 8th to 6th ranking. This indicates that the faculty perceive themselves as digressing less after listening to their tapes, than they originally thought.

The data in Table 2 indicate the number of faculty (total N=19) who, as a function of listening to the tape, shifted:

1. up (+, thus giving a more favorable rating to themselves),
2. down (-, thus giving a more unfavorable rating to themselves),
3. no change.

It is to be noted that five of the ten categories showed a directional change by a majority (≤10/19) of the raters. These categories were:

1. Speaking Voice,
7. Organization and Preparation,
2. Mannerisms or Pleonasms,
5. Enthusiasm Engendered in Class, and
10. General Atmosphere Created.

Table 2 also reveals the ranking of each category

Table 2

Category	Direction: Favorable +	Unfavorable -	No Change 0	Total Changing	Susceptibility to change Rank
1. Speaking voice	8	6	5	14	1.
2. Mannerisms or Pleonasms	4	6	9	10	4.5
3. Knowledge of subject matter	1	1	17	2	10.
4. Enthusiasm - self	2	4	13	6	9.
5. Enthusiasm - engendered in class	6	4	9	10	4.5
6. Digressions	6	1	12	7	7.5
7. Organization and preparation	7	5	7	12	2.
8. Use of analogies, examples, illustrations, etc.	2	6	11	8	6.
9. Handling of questions	3	4	12	7	7.5
10. General atmosphere created	6	5	8	11	3.

Table 2 Directional change by frequency for each category, (N=19), and rank of susceptibility to change of each category.

from most susceptible to least susceptible to change of rating by the faculty as a function of listening to the taped lectures. Category 1, (Speaking Voice) ranked first in susceptibility to change, while category 3 (Knowledge of Subject Matter) was low or least susceptible. Interviews with experienced faculty revealed that they could not acclimate themselves to the sound of their own voice regardless of past auditory exposure. Of interest are the values of category 6 (Digressions), and category 10 (General Atmosphere) with ranks of 7.5 and 3 respectively. This may indicate that faculty are more fearful of digressing than the facts might warrant, and that atmosphere effects continue to be a rather evanescent quality.

Table 3

Number of Faculty who changed ratings preponderantly (>5) upward	= 5
Number of Faculty who changed ratings preponderantly (>5) downward	= 6
Number of Faculty who had no preponderant shift	= 8
TOTAL	19

The data in Table 3 indicate the number of faculty who changed their rating of themselves on the second administration, and the directionality of the change, on five or more of the ten categories. Of the nineteen faculty involved, five rated themselves more favorably the second administration, six less favorably, and eight did not change their ratings on five or more categories. Thus, more than half of the faculty (11/19) appeared sensitive to the in form-

ation obtained from the tape. It is in this sense, then, that the auto-critique method seems of value.

In the more qualitative or second facet of the data analysis, some comments made by the faculty respondents are listed below.

1. The day recorded happened to be a good day! Not every day is like that.

2. After hearing recording of lecture--hard to determine enthusiasm of students--difficult to pick up on tape!

3. Lecture was very boring, I think!

4. I know it, although I often can't articulate it well.

5. The fact that this was a "tape" somewhat cut down on the class response--perhaps I was more formal--but I found this whole thing enlightening!

6. Good questions and enthusiasm following tape (2nd period).

7. Pace could be speeded up.

8. I believe my major fault is in delivery--lack of inflection and too many pauses while trying to phrase thoughts. I think this is offset by inspiring interest through relating subject matter to everyday life.

9. My delivery is too slow--too many pauses--also too many repetitions.

10. I feel that students were more interested in recorder than they were my lecture.

11. Too many long pauses--repeating words--"It is clear", "rather obvious", "we know", "so we see".

Again, these comments seem illustrative of the point that this technique can provide feedback to the interested instructor.

Conclusions

The prime conclusion reached by the investigator is that the auto-critique method is not a useful instrument in

evaluating instruction for a merit criterion purpose. Because of lack of external (to the individual instructor's own frame of reference) criteria, it does not seem to have objective value for inter-instructor or instructor-criterion comparisons. A corollary, if secondary, conclusion, however, is that the auto-critique method has great value for instructional improvement on the part of the average as well as highly motivated instructor. In addition to the anonymity and self pacing features as indicated in the introductory passages of this paper, the results, both quantitative and qualitative, suggest that informational feedback of a differential nature is provided the instructor. Thus, the auto-critique method, as it relates to instructional improvement, seems worthy of continuation and Columbus College has incorporated it into the general orientation program for new faculty during their first year on campus. It is hoped that this method, with additional experimentation, preferably with video tape, will ultimately produce increasingly better teachers.